



CATEGORY:

CLEARED

---

ADDRESS  
CONTACT IF FOUND:

FORM PTO-1390  
(Rev. 10-96)

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER

**TRANSMITTAL LETTER TO THE UNITED STATES  
DESIGNATED/ELECTED OFFICE (DO/EO/US)  
CONCERNING A FILING UNDER 35 U.S.C. 371**

026600-285

U.S. APPLICATION NO. (If known, see 37 C.F.R. 1.5)

097462635

INTERNATIONAL APPLICATION NO.  
PCT/GB98/02043INTERNATIONAL FILING DATE  
13 July 1998PRIORITY DATE CLAIMED  
11 July 1997TITLE OF INVENTION  
CATEGORISING NUCLEIC ACIDAPPLICANT(S) FOR DO/EO/US  
Günter SCHMIDT, Andrew Hugin THOMPSON

It is requested that this Application be prosecuted on the basis of amended Claims 1 to 36 that are attached to the International Preliminary Examination Report and as further amended in the Preliminary Amendment filed herewith.

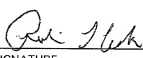
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and the PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
  - a. ☒ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☒ has been transmitted by the International Bureau.
  - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
  - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☐ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11. to 16. below concern other document(s) or information included:

11. ☒ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☒ A FIRST preliminary amendment.  
☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:

Notification of Recording of a Change  
International Preliminary Examination Report with amended claim sheets

U.S. APPLICATION NO. (if known, see 37 C.F.R. 1.50)		INTERNATIONAL APPLICATION NO. PCT/GB98/02043		ATTORNEY'S DOCKET NUMBER 020600-285	
17. <input checked="" type="checkbox"/> The following fees are submitted:				CALCULATIONS	PTO USE ONLY
<b>Basic National Fee (37 CFR 1.492(a)(1)-(5)):</b> Search Report has been prepared by the EPO or JPO ..... \$840.00 (970) International preliminary examination fee paid to USPTO (37 CFR 1.482) ..... \$670.00 (956) No international preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2)) ..... \$690.00 (958) Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO ..... \$970.00 (960) International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4) ..... \$96.00 (962)					
<b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b>				\$ 840.00	
Surcharge of \$130.00 (154) for furnishing the oath or declaration later than months from the earliest claimed priority date (37 CFR 1.492(e)). 20 <input type="checkbox"/> 30 <input type="checkbox"/>				\$ 0.00	
Claims	Number Filed	Number Extra	Rate		
Total Claims	36 -20 =	16	X\$18.00 (966)	\$ 288.00	
Independent Claims	2 -3 =	0	X\$78.00 (964)	\$ 0.00	
Multiple dependent claim(s) (if applicable)			+ \$260.00 (968)	\$ 0.00	
<b>TOTAL OF ABOVE CALCULATIONS =</b>				\$ 1128.00	
Reduction for 1/2 for filing by small entity, if applicable. Verified Small Entity statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28).				\$ 564.00	
<b>SUBTOTAL =</b>				\$ 564.00	
Processing fee of \$130.00 (156) for furnishing the English translation later than months from the earliest claimed priority date (37 CFR 1.492(f)). 20 30 <input type="checkbox"/>				\$ 0.00	
<b>TOTAL NATIONAL FEE =</b>				\$ 564.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 (581) per property +				\$ 0.00	
<b>TOTAL FEES ENCLOSED =</b>				\$ 564.00	
				Amount to be: refunded \$	
				charged \$	
<p><b>NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.</b></p> <p>SEND ALL CORRESPONDENCE TO:</p> <p>Samuel C. Miller, III BURNS, DOANE, SWECKER &amp; MATHIS, L.L.P. P.O. Box 1404 Alexandria, Virginia 22313-1404</p>					
				 SIGNATURE	
				Robin L. Teskin NAME	
				35,030 REGISTRATION NUMBER	

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of )  
)  
Günter SCHMIDT et al. ) Group Art Unit: Unassigned  
)  
Application No. Unassigned ) Examiner: Unassigned  
(Corresponds to PCT/GB98/02043) )  
)  
Filed: January 11, 2000 )  
)  
For: CATEGORISING NUCLEIC )  
ACID )

**PRELIMINARY AMENDMENT**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

Prior to examination, please first amend the above-identified application as follows:

**IN THE CLAIMS**

Claim 9, line 1, please delete "or claim 8".

Claim 10, line 1, please delete "or claim ".

Claim 11, line 1, please delete "any of claims 5 and 8-10" and insert therefor  
--claim 5--.

Claim 12, line 1, please delete "any preceding claim" and insert therefor --claim  
1--.

Claim 13, line 1, please delete "any preceding claim" and insert therefor --claim  
1--.

09/462635-041000

Claim 16, line 1, please delete "or claim 15".

Claim 19, line 1, please delete "any of claims 14-18" and insert therefor --claim  
14--.

Claim 20, line 1, please delete "any of claims 14-18" and insert therefor --claim  
14--.

Claim 23, line 1, please delete "any preceding claim" and insert therefor --claim  
1--.

Claim 25, line 1, please delete "any preceding claim" and insert therefor --claim  
1--.

Claim 29, line 1, please delete "any of claims 26-28" and insert therefor --claim  
26--.

Claim 30, line 1, please delete "any of claims 26-29" and insert therefor --claim  
26--.

Claim 31, line 1, please delete "any of claims 26-29" and insert therefor --claim  
26--.

Claim 32, line 1, please delete "any of claims 26-31" and insert therefor --claim  
26--.

Claim 36, line 1, please delete "any of claims 26-35" and insert therefor --claim  
26--.

**REMARKS**

Favorable consideration on the merits is respectfully requested.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

By: \_\_\_\_\_



Robin L. Teskin

Registration No. 35,030

P.O. Box 1404  
Alexandria, Virginia 22313-1404  
(703) 836-6620  
Date: January 11, 2000

Applicant or Patentee: Günter Schmidt et al.

Application or Patent No.: 09/462,635

Filed or Issued: January 11, 2000

For: CATEGORISING NUCLEIC ACID



**VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS  
(37 C.F.R. §§ 1.9(f) AND 1.27(c)) - SMALL BUSINESS CONCERN**

I hereby declare that I am

- ☐ the owner of the small business concern identified below;  
☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF CONCERN Brax Group Limited

ADDRESS OF CONCERN 13 Station Road

Cambridge CB1 2JB, United Kingdom

I hereby declare that the above-identified small business concern qualifies as a small business concern as defined in 13 C.F.R. § 1.21 for purposes of paying reduced fees under Sections 41(a) and 41(b) of Title 35, United States Code, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average, over the previous fiscal year of the concern, of the persons employed on a full-time, part-time, or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention entitled CATEGORISING NUCLEIC ACID by inventor(s) Günter Schmidt and Andrew Hugin Thompson described in

- ☐ the specification filed herewith  
☒ Application No. 09/462,635, filed January 11, 2000  
☐ Patent No. \_\_\_\_\_, issued \_\_\_\_\_

If the rights held by the above-identified small business concern are not exclusive, each individual, concern, or organization having rights to the invention is listed below,\* and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 C.F.R. § 1.9(c), or by any concern that would not qualify as either a small business concern under 37 C.F.R. § 1.9(d) or a nonprofit organization under 37 C.F.R. § 1.9(e).

\*NOTE: Separate verified statements are required from each named person, concern, or organization having rights to the invention averring to their status as small entities. (37 C.F.R. § 1.27.)

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_  
☐ individual ☐ small business concern ☐ nonprofit organization

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_  
☐ individual ☐ small business concern ☐ nonprofit organization

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earlier of the issue fee and any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 C.F.R. § 1.28(b).)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code; and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING DR. G. H. SCHMIDT

TITLE OF PERSON OTHER THAN OWNER DIRECTOR

ADDRESS OF PERSON SIGNING HOUGHTON MANOE, HOUGHTON, CAMBRIDGE,  
PELT 2BR, UNITED KINGDOM

SIGNATURE [Signature] DATE 2-22-00



CATEGORISING NUCLEIC ACID

The present invention concerns a method for categorising nucleic acid. In particular, the invention concerns a method for sorting nucleic acid, which method permits reduction in the complexity of a nucleic acid population of approximately one order of magnitude, or more. The invention also relates to a kit for carrying out the above method.

Analysis of nucleic acids is fundamental to much of modern molecular biology. A particular feature of nucleic acids derived from living organism is that they are almost invariably complex populations of sequences present in widely varying quantities. In order to characterise these populations of nucleic acids it is usual to attempt to reduce the complexity of the population of nucleic acids in some way. Traditionally the approach has been to clone complex nucleic acid molecules into vectors to allow them to be isolated and either sub-cloned further or analysed directly. Cloning requires the use of biological hosts and these are often difficult to use and require a great deal of specialist knowledge for the cloning procedures to be successful. The traditional processes of cloning to generate libraries of sequences are also only partially automatable.

A problem which cloning does not address is how to isolate sequences which are present only at low copies in backgrounds of sequences present at high copy numbers. Various techniques have been developed to 'normalise' complex nucleic acid populations prior to cloning in order to increase the quantities of sequences at low copy numbers relative to those at high copy numbers. Subtractive hybridisation methods have been used to try and normalise cDNA populations.

PCT/GB93/01452 describes methods of molecular sorting which uses restriction endonucleases that generate ambiguous sticky-ends in the nucleic acid sample to be sorted. Adapters are designed with sticky ends complementary to a single sticky-end sequence or a subset of the these ambiguous sticky ends such that the individual sticky end or subset thereof is coupled to a distinct sequence in the double stranded region of the adapter. This allows subsets of the

adaptored nucleic acid to be amplified using specific primers corresponding to sequences within the adapter which in turn relate to the sequence of the sticky end of the adapter. US patent 5,508,169 (issued November 7, 1995) describes methods very similar to those disclosed in PCT/GB93/01452.

A problem with the above method is that the nucleic acids can be sorted only according to the sequence present on the sticky-ends of the nucleic acid. The sticky-end sequence is of limited length, as determined by the choice of restriction enzyme, thus the basis for sorting is limited.

It is an object of the present invention to provide a method which overcomes the above problems, and provides a wider basis on which sorting of nucleic acid populations can be carried out, not limited by the sticky-end sequence. It is also an object of this invention to provide methods to reduce the complexity of nucleic acid populations by allowing them to be sorted into sub-populations without cloning and to permit normalisation of these populations. This invention describes methods of sorting nucleic acid molecules that have a variety of applications including gene expression profiling, preparation of templates for sequencing, linkage analysis, etc. This invention provides methods of generating sorted libraries. In many applications it is preferable that these sorted nucleic acids be captured on a solid phase support.

Accordingly, the present invention provides a method for categorising nucleic acid, which method comprises producing a nucleic acid population by action of an endonuclease on double-stranded nucleic acid, such that each nucleic acid in the nucleic acid population has a double-stranded portion, contacting the nucleic acid population with one or more oligonucleotide sequences, and isolating nucleic acid which correctly hybridises to an oligonucleotide sequence, wherein each oligonucleotide sequence has a pre-determined recognition sequence, the nucleic acid being categorised by its ability to correctly hybridise to oligonucleotide sequences having the recognition sequence, the recognition sequence being situated such that it recognises a sequence in the double-stranded portion of the nucleic acid,

one or more different recognition sequences being represented in the oligonucleotide sequences.

The present invention also provides kit for categorising a nucleic acid, comprising one or more adaptors and one or more sets of oligonucleotide sequences, wherein the adaptors comprise nucleic acid having a double-stranded primer portion of a known sequence and a single-stranded portion of a pre-determined length, either each single-stranded portion of each nucleic acid in the adaptors having the same pre-determined sequence or all possible sequences of the single-stranded portion being represented in the adaptors, and wherein each oligonucleotide sequence comprises a first sequence, a second sequence attached to the first sequence and a third sequence attached to the second sequence, in which the first sequence is complementary to the sequence of the primer portion of the adaptor, the second sequence is the same sequence as the single-stranded portion of the adaptors or all possible second sequences of the same length as the single-stranded portion of the adaptors are represented within the set of oligonucleotides, and the third sequence comprises a pre-determined recognition sequence.

The invention will now be described in further detail by way of example only, with reference to the accompanying drawings, in which:

Figure 1 shows a schematic of the treatment of a genomic DNA clone with a frequent cutting restriction endonuclease, such as *Sau3A1*, followed by ligation of adaptors to restriction fragments bearing specific primer sequences - all fragments are dealt with simultaneously, but for simplicity only one is shown;

Figure 2 shows a schematic of an amplification step, following the steps of Figure 1, in which fragments are amplified by PCR using adaptor primers;

Figure 3 shows a step following the step of Figure 2, in which amplified fragments are subdivided into 10, wells, each well being identified by a pair of primers used to sort added molecules, each well initially containing one of the pair of primers, there being 4 primers

each with one base probe sequence and each well having 1 of 10 possible pairs generated by a combination of the four primers, the second primer being added after one cycle of synthesis of the first;

Figure 4 shows a schematic of a differential amplification step, following the step of Figure 3, in which the contents of a well containing a primer terminated with AC followed by a probe terminated by AG is amplified and then one cycle of synthesis is performed with the first primer and double strands captured with avidinated beads;

Figure 5 shows a schematic of steps subsequent to those of Figure 4, in which the non-immobilised strand is melted off and washed away and the reaction residue polymerised, a second primer then being added and a second cycle of synthesis performed; and

Figures 6A and 6B show a schematic of steps subsequent to those of Figure 5, in which the non-immobilised strand is melted off and transferred to a fresh reaction vessel, and both primers are then added to the fresh free strand to amplify by PCR.

In the present invention, the nucleic acid population is not isolated (such as by capture onto a solid phase) prior to contacting it with the oligonucleotide sequence(s). Thus each nucleic acid in the population may initially move freely in the suspension or solution in which it is contained. After contacting the nucleic acid population with the oligonucleotide sequence(s), preferably only the nucleic acid(s) which have correctly hybridised to the oligonucleotide sequence(s) are isolated (preferably by capture onto a solid phase).

In more detail, the method of this invention may comprise the following steps:

1. Restricting a large nucleic acid or population of large nucleic acids to generate fragments with known termini.
2. Ligating adaptors or linkers to the termini of these nucleic acid molecules. The ligated adaptor provides a known sequence at the termini of a population of nucleic acids which can be used to design primers which extend beyond the terminal adaptor sequence into unknown sequence adjacent to the known adaptor sequence allowing the unknown sequence to be probed.

3. Optionally amplifying the adapted fragments using primers complementary to the whole or part of the adaptor sequences at the termini of the adapted fragments.
4. Optionally normalising the population of adapted nucleic acids.
5. Selectively amplifying subsets of the nucleic acids through the use of pairs of primers which partially overlap into the unknown sequence. The overlapping primer will hybridise to a subset of the whole population. The size of the subset is determined by the length of overlap of the primer into the adjacent sequence.

The methods of this invention may be applied cyclically to sub-populations of sorted nucleic acids generated by the methods of this invention. Each cycle further reduces the complexity of the population. If necessary the cycles can be repeated until unique nucleic acid is obtained.

In a preferred embodiment the step of restricting nucleic acid is coupled to the ligation of adapters. Preferred restriction endonucleases for use with this invention cleave within their recognition sequence generating sticky-ends that do not encompass the whole recognition sequence. This allows adapters to be designed that bear sticky ends complementary to those generated by the preferred restriction endonuclease but which do not regenerate the recognition site of the preferred restriction endonuclease. This means that if the restriction reaction is performed in the presence of ligase and adapters, the ligation of restriction fragments to each other is reduced by continuous cleavage of these ligations whereas ligation of adapters is irreversible so the presence of adapters drives the restriction to completion and similarly the restriction endonuclease drives the ligation reaction to completion. This process ensures that a very high proportion of restriction fragments are ligated to adapters. This is advantageous as ligation of adapters to restriction fragments is a relatively inefficient process. This is due to random ligation of restriction products to each other if these are phosphorylated. In this embodiment the adapters used are preferably not phosphorylated at their 5' hydroxyl groups so that they cannot ligate to themselves.

GB 9115407.0 describes a method of normalising a population of nucleic acids comprising the following steps:

1. Combining a mixture of heterogeneous DNA fragments with oligonucleotide primers compatible with some nucleic acid amplification system and denaturing the double stranded heterogeneous DNA.
2. Altering the conditions, i.e. reducing the temperature, to allow the more common species to re-anneal while preventing the primers from annealing to the DNA. The temperature for re-annealing at this stage must be higher than the melting temperature of the PCR primers.
3. Altering the reaction conditions further to allow the PCR primers to anneal to the remaining single stranded DNA which should represent the rarer species.
4. Performing strand extension of the primed species.

Advantageously, the above steps are applied cyclically a number of times to amplify the rarer species to a significant extent.

Application of this method to sequences with known termini permits the design of primers with very specific melting temperatures allowing the method to be used generically. Use of this method is particularly advantageous in reducing the complexity of genomic DNA as a significant proportion of most genomic DNA is repetitive sequence.

The advantage of providing a known sequence adjacent to probe sequence allows one to design libraries of probes, where all the probes in a library have the same melting temperature. This is advantageous as hybridisation of the entire library can be performed simultaneously at a single temperature whilst retaining the stringency of hybridisation.

Consider a large DNA fragment such as a mitochondrial genome or a cosmid or a microbial genome. To perform steps 1 to 4 of the method described above, such a large molecule can be cleaved with a frequently cutting restriction enzyme to generate fragments of the order of a few hundred bases in length. If a restriction endonuclease like Sau3A1 is used fragments with a

known sticky end are left, to which double stranded adaptors can be ligated. These adaptors will bear a known primer sequence, and a sticky end complementary to that produced by the restriction endonuclease to permit ligation. A combined restriction and ligation protocol as described above is appropriate.

The majority of properly restricted fragments as a result bear an adaptor at each of their termini. This permits amplification of the adaptored restriction fragments at this stage if that is desired. After adaptoring and any non-selective amplification and normalisation, the nucleic acids can be differentially amplified to generate specific subsets of the starting population. The method of differential amplification preferably comprises the following steps:

1. Dividing the adaptored population of restriction fragments into separate wells. If, for example, primers with an overlap of a single base are used then the adaptored fragments would be divided into 10 or 16 wells.
2. Adding to each well one type of biotinylated primer of a predetermined set. The primer bears a sequence complementary to that provided by the adaptor and restriction site. The primer additionally bears an overlap of a predetermined number of bases beyond the known sequence into the unknown sequence immediately adjacent to the restriction site. Primers with different overlaps are added to different well. Four primers are need if a 1 base overlap is used. If 16 wells are used each of the 4 primers are added to 4 wells.
3. Denaturing the amplified fragment population that was subdivided into each well by raising the temperature. The temperature is then reduced to permit the primer sequences to anneal. Primers preferably have equalised melting temperatures so that conditions for use of all primers are the same.
4. Adding thermostable polymerase and nucleotides to extend annealed primers.
5. Capturing the biotinylated strand extension products from (4) onto a solid phase substrate derivitised with avidin. This may be effected through the addition of avidinated beads. These may optionally be magnetic beads.
6. Melting off the non-biotinylated complementary strand and washing this away. This leaves a single stranded copy of the selected fragments immobilised on the solid phase support.

7. To each of the separate pools is added one of the same set of primers as used in step (2) but not biotinylated, such that each pool receives a different combination of primers from this step and step (2). The primers should anneal to the single stranded capture molecules from (6). If 16 pools are used, to each is added one of the same 4 primers, but not biotinylated such that each of the 16 pools carries one of the possible different combinations of pairs of the 4 primers.
8. Extending the primed captured strands with polymerase and nucleotide triphosphates.
9. Denaturing the free strand from the captured strand by raising the temperature. The 'selected' free strand is thus released into solution. The liquid phase can be transferred to fresh reaction vessel or the solid phase support bearing the captured strands from (5) can be removed. This is very easy if the support used are magnetic beads as these can be removed by electromagnetic attraction to a probe.

The isolated free strands from (9) are thus isolated. At this stage the selected strands can be captured onto a solid phase support or amplified or the process of differential amplification can be repeated on the isolated subsets generated to further sub-sort these populations. This would be effected by using primers which overlap further into the unknown sequence adjacent to the known sequence of the adapter and the selected fragment. The sorted fragment could equally be cloned into a biological vector at this stage if desired.

Generating a captured library is advantageous in that it facilitates easy manipulation of the library of fragments. Such manipulations include copying, amplification and probing of the library for particular sequences. A captured library dispenses with any requirement for biological cloning vectors to maintain the library as such a library can be readily copied using polymerases and nucleotide triphosphates. The captured library can be readily washed and can very easily be stored in a refrigerated environment.

It should be noted in the example of primers that overlap by a single base, that the amplification products from the well containing a primer terminated by A followed by the primer terminated by G gives the complement of the well where G is followed by A. It might therefore be



desirable to pool the reactions of where the same pair of primers are present but used in a different order to ensure that both strands of each DNA molecule are present and captured on the solid phase support. This would thus give 10 different pools. This is a convenient number as one can reduce the complexity of a library by one order of magnitude with four primers. Each sorted library of fragments can be further sub-sorted to an arbitrary degree.

An alternative embodiment of this method uses primers already immobilised on a solid phase support, preferably covalently linked to the support instead of biotinylated primers in step (2) of the differential amplification process. Such solid phase supports can be magnetic beads, as described in EP-A-0 091 453 and EP-A-0 106 873, or the support could be polymer beads. PCT/GB92/02394 describes a solid phase polymer support in a micro-column where the solid phase support are beads of silica gel. The beads are retained between two frits in the column through which solvents and reagents can flow. Such apparatus is also applicable with this invention.

One can clearly repeat the sorting process starting from a captured library that has been previously sorted.

One can also clearly use just 10 wells to generate sorted populations as all of the sequence information in a series of 16 wells will be present if just the 10 different pairs of primer combinations are used.

It should also be clear that labels can be introduced into sorted molecules by the primers used as part of the sorting process. Methods of introducing labels into primer oligonucleotides are well known in the art. Biotin has been discussed above, but many others are applicable.

One can also use probes which overlap beyond the provided adaptor sequence to any extent. It becomes more difficult, however, to ensure the stringency of hybridisation as the number of bases extending into the unknown sequence from the adaptor is increased.

000110-5899900

To effect higher degrees of sorting one can either sort a sorted library with a set of four primers that overlap beyond the known terminal sequences by a single base or one can use primers with a longer sequence overlap. To sort an adapted population of nucleic acid fragments using primers with a 2 base overlap beyond the adaptor sequence, the adapted population of restriction fragments is sub-divided into 256 wells. In each well is one of 16 biotinylated primers which bear a sequence complementary to that provided by the adaptor and restriction site. The primers additionally bear an overlap of 2 bases beyond the known sequence into the unknown sequence immediately adjacent to the restriction site. The amplified fragment population subdivided into each well is denatured by raising the temperature and cooled allowing the primer sequences to anneal. Primers, again, preferably have equalised melting temperatures so that conditions for use of all primers is the same. Thermostable polymerase and nucleotides are added to extend annealed primers. Biotinylated fragments are captured onto a solid phase substrate via avidin and the complementary strand is melted off and washed away. To each of the 256 pools is added one of the same 16 primers, but not biotinylated such that each of the 256 pools carries one of the possible different combinations of pairs of the 16 primers. Again, AT followed by GC gives the complement of the reaction of GC followed by AT so it might be desirable to pool these pairs to give a total of 136 pools. For an overlap of  $n$  bases, one can distinguish  $4^n$  distinct sequences. If both termini of a molecule are used to select fragments then one can distinguish fragments into  $(4^n \times (4^n + 1)/2)$  distinct sets, since the orientation of each fragment is unknown.

Sorting a library resolves fragments from a large, complex population into defined sets whose size will be statistically regular and determinable as long as the size of the parent library is known, even if only approximately. The composition of the sorted library will be less complex than that of the parent library. This allows for useful manipulations of a large library without loss of information as all the sequences present in the starting library should be present in one of the sub libraries as long as all of the possible sub-libraries are generated. This

method offers greater ease of manipulation of complex nucleic acid libraries and greater precision of manipulation than cloning into biological vectors.

To put this invention into practise requires the construction of probe oligonucleotides (ONs). Precise control over hybridisation conditions will be required to ensure clean results in differential amplification.

Details and reviews on the construction of labelled and modified ONs are available in numerous up-to-date texts, see references 1 to 6 below. A brief discussion of preferred design possibilities is given below.

There are major differences between the stability of short oligonucleotide duplexes containing all Watson-Crick base pairs. For example, duplexes comprising only adenine and thymine are unstable relative to duplexes of guanine and cytosine only. These differences in stability can present problems when trying to hybridise mixtures of short oligonucleotides to a target RNA. Low temperatures are needed to hybridise A-T rich sequences but at these temperatures G-C rich sequences will hybridise to sequences that are not fully complementary. This means that some mismatches may occur, and specificity can be lost for the G-C rich sequences. At higher temperatures G-C rich sequences will hybridise specifically but A-T rich sequences will not hybridise.

It is desirable that probes within a library behave in a similar manner, i.e. they should have similar melting temperatures and preferably also binding kinetics. In order to normalise these effects, modifications can be made to nucleic acids. Modifications fall into three broad categories: base modifications, backbone modifications and sugar modifications.

#### *Base modifications*

Numerous modifications can be made to the standard Watson-Crick bases. The following are examples of modifications that should normalise base pairing energies to some extent but they are not limiting:

- The adenine analogue 2,6-diaminopurine forms three hydrogen bonds to thymine rather than two and therefore forms more stable base pairs.
- The thymine analogue 5-propynyl dU forms more stable base pairs with adenine.
- The guanine analogue hypoxanthine forms two hydrogen bonds with cytosine rather than three and therefore forms less stable base pairs.

These and other possible modifications should make it possible to compress the temperature range at which short oligonucleotides can hybridise specifically to their complementary sequences.

#### *Backbone modifications*

Nucleotides may be readily modified in the phosphate moiety. Under certain conditions, such as low salt concentration, analogues such as methylphosphonates, triesters and phosphoramidates have been shown to increase duplex stability. Such modifications may also have increased nuclease resistance. Further phosphate modifications include phosphodithirates and boranophosphates, each of which increase the stability of ONs.

Isosteric replacement of phosphorus by sulphur gives nuclease resistant ONs (reference 7). Replacement by carbon at either phosphorus or linking oxygen is also a further possibility.

#### *Sugar modifications*

Various modifications to the 2' position in the sugar moiety may be made (references 12 and 13). The sugar may be replaced by a different sugar such as hexose or the entire sugar phosphate backbone can be entirely replaced by a novel structure such as in peptide nucleic acids (PNA). For a discussion see reference 8. PNA may be the ideal choice as it forms duplexes of the highest thermal stability of any analogues so far discovered.

One major source of error in hybridisation reactions is the stringency of hybridisation of the primers to the target sequence and to the unknown bases beyond. If the primers designed for a target bear single artificially introduced mismatches the discrimination of the system is much higher (Zhen Guo *et al.*, Nature Biotechnology 15, 331-335, April 1997). Additional mismatches are not tolerated to the same extent that a single mismatch would be when a fully complementary primer is used. Thus this can be exploited in the method disclosed above. If the probe used to extends beyond the provided sequence by 1 base, an artificial mismatch, 1 helical turn away from the probe base destabilises the double helix to a considerable degree if there is a second mismatch at the probe site.

Details on effects of hybridisation conditions for nucleic acid probes can be found in references 9 to 11.

Mass labels for use in the present invention are disclosed in patent application PCT/GB98/00127. Further labels for use in the present invention are discussed in the UK applications of Page White & Farrer file numbers 87820, 87821, 87900.

## References

- (1) Gait, M.J. editor, 'Oligonucleotide Synthesis: A Practical Approach', IRL Press. Oxford. 1990
- (2) Eckstein. editor, 'Oligonucleotides and Analogues: A Practical Approach', IRL Press. Oxford, 1991
- (3) Kricka, editor. 'Nonisotropic DNA Probe Techniques', Academic Press, San Diego, 1992
- (4) Haugland, 'Handbook of Fluorescent Probes and Research Chemicals', Molecular Probes. Inc., Eugene, 1992
- (5) Keller and Manack, 'DNA Probes, 2nd Edition', Stockton Press, New York, 1993
- (6) Kessler, editor, 'Nonradioactive Labelling and Detection of Biomolecules', Springer-Verlag, Berlin, 1992.
- (7) J.F. Milligan. M.D. Matteucci, J.C. Martin, J. Med. Chem. 36(14), 1923 - 1937, 1993.
- (8) P.E. Nielsen, Annu. Rev. Biophys. Biomol. Struct. 24, 167 - 183, 1995.
- (9) Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26, 227-259, 1991
- (10) Sambrook et al. 'Molecular Cloning: A Laboratory Manual, 2nd Edition', Cold Spring Harbour Laboratory, New York, 1989
- (11) Hames, B.D., Higgins, S.J., 'Nucleic Acid Hybridisation: A Practical Approach', IRL Press, Oxford, 1988
- (12) C.J. Guinosso, G.D. Hoke, S.M. Freier, J.F. Martin, D.J. Ecker, C.K. Mirabelle, S.T. Crooke, P.D. Cook, Nucleosides Nucleotides 10, 259 - 262, 1991.
- (13) M. Carmo-Fonseca, R. Pepperkok, B.S. Sproat, W. Ansorge, M.S. Swanson, A.I. Lamond, EMBO J. 7, 1863 - 1873, 1991.

**Claims:**

1. A method for categorising nucleic acid, which method comprises producing a nucleic acid population by action of an endonuclease on double-stranded nucleic acid, such that each nucleic acid in the nucleic acid population has a double-stranded portion, contacting the nucleic acid population with one or more oligonucleotide sequences, and isolating nucleic acid which correctly hybridises to an oligonucleotide sequence by capturing the oligonucleotide sequence on a solid phase, wherein each oligonucleotide sequence has a pre-determined recognition sequence, the nucleic acid being categorised by its ability to correctly hybridise to oligonucleotide sequences having the recognition sequence, the recognition sequence being situated such that it recognises a sequence in the double-stranded portion of the nucleic acid, one or more different recognition sequences being represented in the oligonucleotide sequences.
2. A method according to claim 1, wherein the endonuclease is selected such that each nucleic acid in the nucleic acid population has a sticky end of a known common length extending from a terminal of its double-stranded portion.
3. A method according to claim 1, wherein the endonuclease is selected such that each sticky end of each nucleic acid in the nucleic acid population has the same known base sequence.
4. A method according to claim 3, wherein prior to contacting the nucleic acid population with the oligonucleotide sequences, the nucleic acid population is contacted with an adaptor to ligate the adaptor to a terminal of each nucleic acid in the nucleic acid population, wherein the adaptor comprises a double-stranded primer portion having a known base sequence, and a single-stranded portion complementary to the known sticky end of the nucleic acids in the nucleic acid population.
5. A method according to claim 4, wherein each oligonucleotide sequence comprises a first sequence, a second sequence attached to the first sequence and a third sequence attached to the second sequence, in which the first sequence is complementary to the sequence of the primer portion of the adaptor, the second sequence is complementary to the known sticky end

of the nucleic acids in the nucleic acid population, and the third sequence comprises the pre-determined recognition sequence.

6. A method according to claim 2, wherein the endonuclease is selected such that the sticky ends of the nucleic acids in the nucleic acid population have a plurality of different base sequences.

7. A method according to claim 6, wherein prior to contacting the nucleic acid population with the oligonucleotide sequences, the nucleic acid population is contacted with an array of adaptors to ligate an adaptor to a terminal of the nucleic acids in the nucleic acid population, wherein each adaptor comprises a double-stranded primer portion having a known base sequence, and a single-stranded portion of the same length as the sticky ends of the nucleic acids in the nucleic acid population, all of the possible base sequences of the single-stranded portion of the adaptor being represented in the array of adaptors.

8. A method according to claim 7, wherein each oligonucleotide sequence comprises a first sequence, a second sequence attached to the first sequence and a third sequence attached to the second sequence, in which the first sequence is complementary to the sequence of the primer portion of the adaptors, the second sequence is of the same length as the sticky ends of the nucleic acids in the nucleic acid population, and the third sequence comprises the pre-determined recognition sequence, and wherein in any one group of oligonucleotides having the same recognition sequence all of the possible base sequences of the second sequence are represented.

9. A method according to claim 5 or claim 8, wherein the recognition sequence consists of one base.

10. A method according to claim 5 or claim 8, wherein the recognition sequence consists of two or more bases.

11. A method according to any of claims 5 and 8-10, wherein in any one group of oligonucleotides having the same recognition sequence the third sequence consists of the



recognition sequence and a pre-determined number of bases situated between the second sequence and the recognition sequence, all possible sequences of the pre-determined number of bases in the third sequence being represented in that group of oligonucleotides.

12. A method according to any preceding claim, wherein the nucleic acid population is amplified by PCR prior to reaction with the oligonucleotide sequences.

13. A method according to any preceding claim, wherein those nucleic acids are isolated both terminals of which correctly hybridise to an oligonucleotide sequence.

14. A method according to claim 13, wherein a first set of oligonucleotide sequences is contacted with the nucleic acid population in a first step by denaturing the nucleic acid population in the presence of the first set of sequences to produce a single-stranded nucleic acid population and allowing the single-stranded nucleic acid to hybridise to the first sequences, immobilising those nucleic acids which correctly hybridise to the first sequences, extending the correctly hybridised oligonucleotide sequences along the single-stranded portion of the immobilised nucleic acid to form double-stranded nucleic acid, denaturing the double-stranded nucleic acid and removing non-immobilised species to isolate the resulting immobilised single-stranded nucleic acid, contacting the immobilised single-stranded nucleic acid with a second set of oligonucleotide sequences in a second step, extending the correctly hybridised oligonucleotide sequences along the immobilised single-stranded nucleic acid to form double-stranded nucleic acid, denaturing the double-stranded nucleic acid and isolating the resulting non-immobilised single-stranded nucleic acid.

15. A method according to claim 14, wherein the extended and isolated products of the first step and/or the extended and isolated products of the second step are amplified by PCR.

16. A method according to claim 14 or claim 15, wherein the correctly hybridised nucleic acids are immobilised by immobilising the oligonucleotide sequences.

17. A method according to claim 16, wherein each oligonucleotide in the first set of sequences carries a biotin residue such that prior to or after hybridising to the nucleic acid the sequence is captured on an avidinated solid phase.

18. A method according to claim 16, wherein each oligonucleotide in the first set of sequences is covalently attached to a solid support prior to contacting with the nucleic acid population.

19. A method according to any of claims 14-18, wherein the recognition sequence of the first and second set of oligonucleotide sequences consists of one base and, prior to performing the first step, the nucleic acid population is sub-divided into 16 wells, each well containing oligonucleotides from the first set of sequences having one of the four possible recognition sequences, and wherein in the second step oligonucleotides from the second set of sequences are added to each well, such that all possible combinations of the identities of the first and second set of oligonucleotide sequences and their order of addition to the well are represented in the 16 wells.

20. A method according to any of claims 14-18, wherein the recognition sequence of the first and second set of oligonucleotide sequences consists of two bases and, prior to performing the first step, the nucleic acid population is sub-divided into 256 wells, each well containing oligonucleotides from the first set of sequences having one of the 16 possible recognition sequences, and wherein in the second reaction oligonucleotides from the second set of sequences are added to each well, such that all possible combinations of the identity of the first and second set of oligonucleotide sequences and their order of addition to the well are represented in the 256 wells.

21. A method according to claim 19, wherein the contents of each pair of wells to which the same pair of oligonucleotide sequences were added but in a different order, are combined to give 10 different wells.

22. A method according to claim 20, wherein the contents of each pair of wells to which the same pair of oligonucleotide sequences were added but in a different order, are combined to give 136 different wells.

AMENDED SHEET

23. A method according to any preceding claim, wherein the oligonucleotide sequences have equalised melting temperatures.

24. A method according to claim 23, wherein the melting temperatures are equalised by incorporating one or more analogues of natural nucleotides into the oligonucleotide sequences, the analogues comprising base modifications, sugar modifications and/or backbone modifications.

25. A method according to any preceding claim, wherein the endonuclease is selected such that it cuts the nucleic acid at a site within the recognition site of the endonuclease.

26. A kit for categorising a nucleic acid, comprising one or more adaptors and one or more sets of oligonucleotide sequences, wherein the adaptors comprise nucleic acid having a double-stranded primer portion of a known sequence and a single-stranded portion of a pre-determined length, either each single-stranded portion of each nucleic acid in the adaptors having the same pre-determined sequence or all possible sequences of the single-stranded portion being represented in the adaptors, and wherein each oligonucleotide sequence comprises a first sequence, a second sequence attached to the first sequence and a third sequence attached to the second sequence, in which the first sequence is complementary to the sequence of the primer portion of the adaptor, the second sequence is the same sequence as the single-stranded portion of the adaptors or all possible second sequences of the same length as the single-stranded portion of the adaptors are represented within the set of oligonucleotides, and the third sequence comprises a pre-determined recognition sequence.

27. A kit according to claim 26, wherein the recognition sequence consists of one base.

28. A kit according to claim 26, wherein in the recognition sequence consists of two or more bases.

29. A kit according to any of claims 26-28, wherein in any one group of oligonucleotides having the same recognition sequence, the third sequence consists of the recognition sequence and a pre-

determined number of bases situated between the second sequence and the recognition sequence, all of the possible sequences of the pre-determined number of bases in the third sequence being represented in that group of oligonucleotides.

30. A kit according to any of claims 26-29, comprising two sets of oligonucleotide sequences, each of the oligonucleotides in one set being biotinylated.

31. A kit according to any of claims 26-29, comprising two sets of oligonucleotide sequences, each of the oligonucleotides in one set being covalently attached to a solid support.

32. A kit according to any of claims 26-31, additionally comprising an endonuclease.

33. A kit according to claim 32, wherein the endonuclease is selected such that when it is reacted with double-stranded nucleic acid, nucleic acids are produced each of which comprises a double-stranded portion.

34. A kit according to claim 33, wherein the endonuclease is selected such that the nucleic acids produced have a sticky end of a known common length extending from a terminal of the double-stranded portion, and wherein each sticky end of each nucleic acid in the nucleic acid population has the same known base sequence.

35. A kit according to claim 33, wherein the endonuclease is selected such that the nucleic acids produced have a sticky end of a known common length extending from a terminal of the double-stranded portion, and wherein the sticky ends of the nucleic acids in the nucleic acid population exhibit a plurality of different base sequences.

36. A kit according to any of claims 26-35, wherein the endonuclease is selected such that it cuts the nucleic acid at a site within the recognition site of the endonuclease.

Figure 1. The structure of the proposed model.

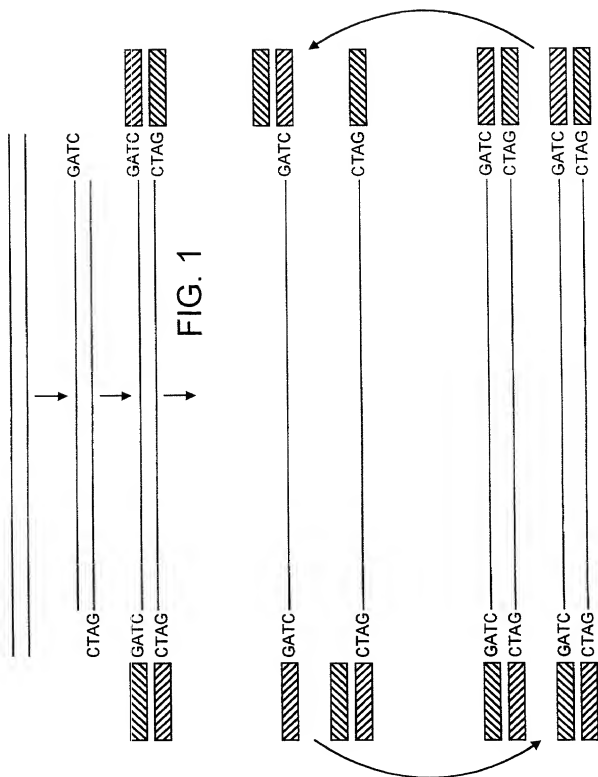


FIG. 2

2 / 5

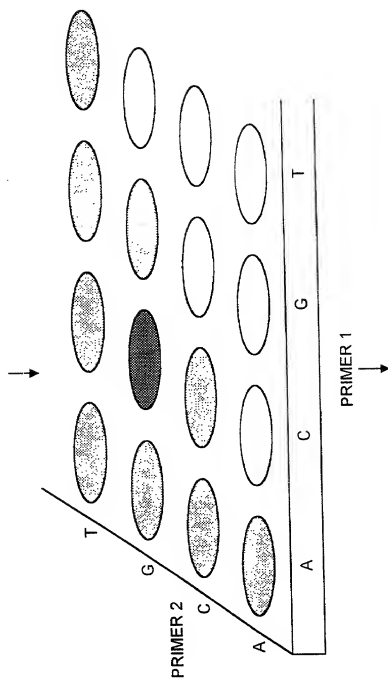


FIG. 3

3 / 5

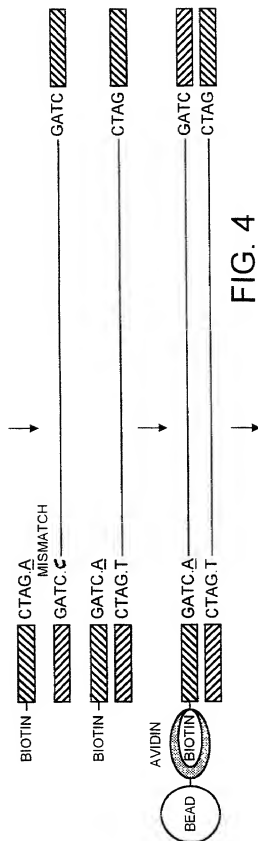


FIG. 4

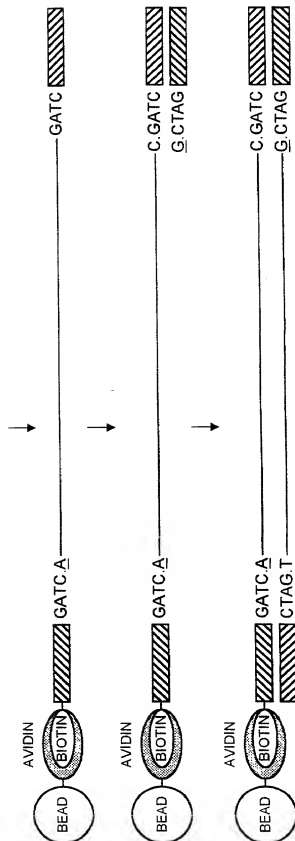


FIG. 5

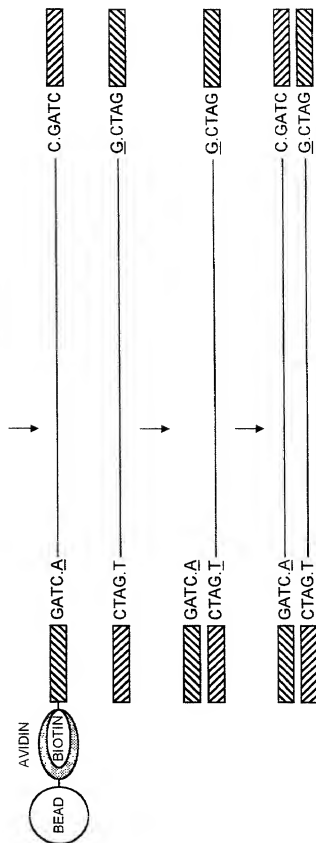


FIG. 6A



5 / 5

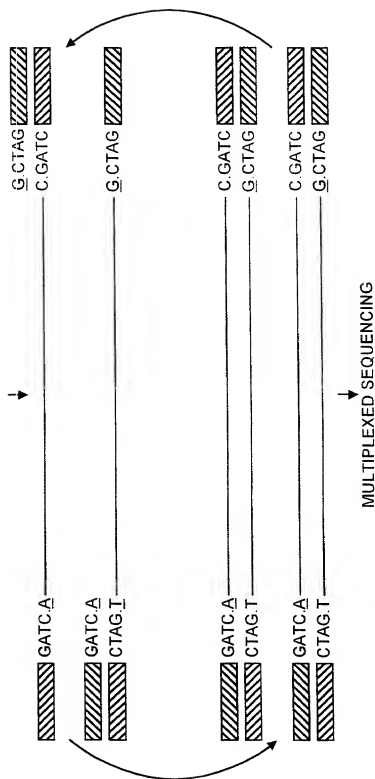


FIG. 6B

**COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY**  
(Includes Reference to Provisional and PCT International Applications)

Attorney's Docket No.  
020600-285

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

CATEGORISING NUCLEIC ACID

the specification of which (check only one item below):

☐ is attached hereto.

☒ was filed as United States application

Number 09/462,635

on January 11, 2000

and was amended

on \_\_\_\_\_ (if applicable).

☒ was filed as PCT international application

Number PCT/GB98/02043

on July 13, 1998

and was amended

on \_\_\_\_\_ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 (a)-(e) of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

**PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. §119:**

COUNTRY (if PCT, indicate "PCT")	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED UNDER 35 U.S.C. §119
United Kingdom	9714716.9	11 July 1997	<u>X</u> Yes _ No
United Kingdom	9719284.3	10 September 1997	<u>X</u> Yes _ No
United Kingdom	9726949.2	19 December 1997	<u>X</u> Yes _ No

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below.

\_\_\_\_\_  
(Application Number)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Application Number)

\_\_\_\_\_  
(Filing Date)



**COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (CONT'D)**  
(Includes Reference to Provisional and PCT International Applications)

Attorney's Docket No.

020600-285

FULL NAME OF SOLE OR FIRST INVENTOR

SIGNATURE

DATE

Günther SCHMIDT

RESIDENCE

Houghton Manor, Houghton, Cambridge, PE17 2BQ, United Kingdom

CITIZENSHIP

German

POST OFFICE ADDRESS

Houghton Manor, Houghton, Cambridge, PE17 2BQ, United Kingdom

FULL NAME OF SECOND JOINT INVENTOR, IF ANY

SIGNATURE

DATE

Andrew Hugin Thompson

RESIDENCE

25 Knoll Park, Alloway, Ayr KA7 4RH, United Kingdom

CITIZENSHIP

United Kingdom

POST OFFICE ADDRESS

25 Knoll Park, Alloway, Ayr KA7 4RH, United Kingdom

FULL NAME OF THIRD JOINT INVENTOR, IF ANY

SIGNATURE

DATE

RESIDENCE

CITIZENSHIP

POST OFFICE ADDRESS

FULL NAME OF FOURTH JOINT INVENTOR, IF ANY

SIGNATURE

DATE

RESIDENCE

CITIZENSHIP

POST OFFICE ADDRESS

FULL NAME OF FIFTH JOINT INVENTOR, IF ANY

SIGNATURE

DATE

RESIDENCE

CITIZENSHIP

POST OFFICE ADDRESS

FULL NAME OF SIXTH JOINT INVENTOR, IF ANY

SIGNATURE

DATE

RESIDENCE

CITIZENSHIP

POST OFFICE ADDRESS

FULL NAME OF SEVENTH JOINT INVENTOR, IF ANY

SIGNATURE

DATE

RESIDENCE

CITIZENSHIP

POST OFFICE ADDRESS

FULL NAME OF EIGHTH JOINT INVENTOR, IF ANY

SIGNATURE

DATE

RESIDENCE

CITIZENSHIP

POST OFFICE ADDRESS